Chapter 49

Discovering Opinions from Customers’ Unstructured Textual Reviews Written in Different Natural Languages

Jan Žižka
Mendel University in Brno, Czech Republic

František Dařena
Mendel University in Brno, Czech Republic

ABSTRACT

Gaining new and keeping existing clients or customers can be well-supported by creating and monitoring feedbacks: “Are the customers satisfied? Can we improve our services?” One of possible feedbacks is allowing the customers to freely write their reviews using a simple textual form. The more reviews that are available, the better knowledge can be acquired and applied to improving the service. However, very large data generated by collecting the reviews has to be processed automatically as humans usually cannot manage it within an acceptable time. The main question is “Can a computer reveal an opinion core hidden in text reviews?” It is a challenging task because the text is written in a natural language. This chapter presents a method based on the automatic extraction of expressions that are significant for specifying a review attitude to a given topic. The significant expressions are composed using significant words revealed in the documents. The significant words are selected by a decision-tree generator based on the entropy minimization. Words included in branches represent kernels of the significant expressions. The full expressions are composed of the significant words and words surrounding them in the original documents. The results are here demonstrated using large real-world multilingual data representing customers’ opinions concerning hotel accommodation booked on-line, and Internet shopping. Knowledge discovered in the reviews may subsequently serve for various marketing tasks.

DOI: 10.4018/978-1-4666-6543-9.ch049
INTRODUCTION

In the modern electronic era, it has become quite common for various organizations, enterprises, companies, and corporations to offer their services, products, and business to customers using the widespread available common world-wide-web tools like the Internet, or similar public or private (intranet) network-based tools (See for example Liu, 2006). From the other side, customers who may take advantage of these modern, effective, and comfortable possibilities can also employ the same means for expressing their open opinions, sentiments, recommendations, and so like, concerning the services or products. These opinions may represent a very valuable information source for respective organizations and enterprises in the form of a very beneficial feedback. Having a sufficient volume of data containing this information, people responsible for the quality of services and products can reveal what went well and what badly, which points were weak and which strong, when and why the customers were satisfied, dissatisfied, or maybe indifferent, and after careful processing of the data improve the business and, therefore, competitiveness. Such improvements ordinarily belong among the typical important tasks of marketing departments in companies and organizations: The relationships with customers should be as good as possible because the competition in the market economy is (and undoubtedly always will be) very hard. Today’s developing areas like, for example, business intelligence focus on a very deep analysis of data, inevitably including – if possible – data coming from the customer side.

In this chapter, the authors concern specifically with very large electronic data deposited in enterprise databases or data warehouses – much more information can be found in (Shmueli et al., 2010). The data may have different forms, for example, prepared questionnaires that allow ticking off suitable answers, filling in one of offered possibilities, or even writing quite freely, in any natural language, and without any forced structure, customers’ review. The rest of this chapter is devoted to the problem how it would be possible to automatically search for customers’ opinions and attitudes hidden in many unstructured textual files. Successful marketing depends, among other things, first and foremost on good processing of as much data as possible because such data includes the information that gives evidence of what takes place. Therefore, an enterprise typically offers a possibility to enter a textual review expressed freely in a natural language to its customers; the reviews are usually limited to a reasonable size by a maximal number of characters. In addition, those reviews are often publicly available, thus other, potentially future customers can read experience of their forerunners and according to the published opinions they can decide whether to accept or decline an offered service or product, or what other alternatives could be utilized. In this way, the marketing department can (and should) collect a lot of very valuable data, especially if it runs for a long time, months or years.

A good and probably well-known example of this method can be found, for instance, on Amazon Web pages (www.amazon.com and www.amazon.co.uk) devoted to various products where customers may write their review and give one (the worst evaluation) to five (the best one) stars. The company publishes those evaluations and opinions, enabling seeing the worst and best ones, or everything. The customers can also add a sign whether a published specific opinion was helpful for them, which can later attract more interested persons without regard to a positive or negative meaning of such a review – both classifications can naturally be useful. Because the customers as authors of reviews can evaluate a product or service by one to five stars, a reader can also expect opinions bringing mixed views – descriptions of both positive and negative properties; and, maybe, for someone, the negative properties do not play a significant role while for another one it does.